

Title	HITL/SITL型AIガバナンスの規制シグナルとイノベーション効果
Author(s)	吉村, 直泰
Citation	年次学術大会講演要旨集, 40: 673-678
Issue Date	2025-11-08
Type	Conference Paper
Text version	publisher
URL	<a href="https://hdl.handle.net/10119/20115">https://hdl.handle.net/10119/20115</a>
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

## 2 B 2 4

# HITL/SITL 型 AI ガバナンスの規制シグナルとイノベーション効果

吉村 直泰（経済産業省・政策研究大学院大学）

## 1. はじめに

AI の進化に伴い、リスク管理とイノベーションを両立させる制度設計が国際的な政策課題となっている。本研究は、自動運転技術を対象に、人間が最終責任を負う Human-in-the-Loop (HITL) と、社会規範を AI に内在化させる Society-in-the-Loop (SITL) の 2 つのガバナンス枠組みが、企業のイノベーション活動に与える影響を比較分析する。責任分担の明確化や利用者信頼の醸成を通じた制度の「需要創出シグナル」に着目し、ポーター仮説を拡張する理論枠組を提示する。企業別の特許データに基づく実証により、規制がイノベーションに正の影響を与える可能性を示す。特に、利用者の責任受容度が低い技術分野では、段階的な SITL 移行が社会受容と技術革新の両立を促す制度設計となることを示唆する。

## 2. 先行研究とその展開

### 2.1. ポーター仮説とその検証

1991 年に Porter がいわゆる「ポーター仮説」と呼ばれる、「適切に設計された環境規制は、費用低減・品質向上につながる技術革新を刺激し、その結果国内企業は国際市場において競争上の優位を獲得し、他方で産業の生産性も向上する可能性がある」との主張を行って以来、主に環境規制の分野で規制とイノベーションの関係について世界各国の企業データやケースを用いた検証が行われてきた。Porter & Van der Linde (1995) は具体的ケースを通じて、環境規制が企業に対する情報シグナルや投資の不確実性減少等を通じて製品・工程の改善を促すことを示し、Ambec, Cohen, Elgie & Lanoie (2013) は、20 年間の実証分析論文をレビューすることで、ポーター仮説の「弱い」バージョン(より厳格な環境規制がより多くのイノベーションにつながる)の証拠はかなり明確で十分に確立されていること、ポーター仮説の「強い」バージョン(規制が厳格化すると業績が向上する)に関する経験的証拠は必ずしも確立されていないが、ポーター仮説の「狭い」バージョン(命令的規制より成果基準や市場型手法の方がイノベーション誘因が強い)は妥当であり、環境保護と競争力は固定的トレードオフではないことを示した。

特に、先行文献でポーター仮説の理論的背景として強調されていることは以下の 3 点である。

- ① リスク回避的で現在バイアス・限定合理性に縛られている企業に対する「規制のシグナル・誘導効果」
- ② 戦略的な規制導入により先行者優位や製品の差別化をもたらすこと、また知識のスピルオーバー効果による過小投資均衡から投資拡大均衡への転換など市場の失敗を規制が緩和する可能性があること
- ③ 新技術導入を阻害していた社内の情報非対称性やガバナンス不全等の組織慣性を規制が打破すること

### 2.2. ポーター仮説の AI ガバナンスとイノベーションの関係への展開可能性

ポーター仮説は主として環境規制の分野で実証分析が積み重ねられてきているが、AI を含む情報技術の分野においても規制がイノベーションを促進する効果がありうる。Blind, Petersen, & Riillo (2017) は、不確実性の高い市場における標準化と規制がイノベーションに与える影響について、特に、企業が市場参入や投資判断を行う際に直面する不確実性を、標準や規制がどのように軽減し、技術革新の方向性に影響を与えるのかを分析しており、AI など技術の将来予測が難しく投資の不確実性の高い分野における規制の役割を示している。特に、ネットワーク外部性や相互運用性が重要な産業では標準化が企業の技術選択を容易にする効果があるとしており、また規制が安全性に関する最低基準を設定することでイノベーションを促進する効果を持つとしており、企業に対するシグナル効果や市場の失敗の緩和効果があることを強調するポーター仮説と通底している。ポーター仮説を環境規制の分野を超えて規制一般にも打倒しうる理論的枠組みととらえ、AI ガバナンスとイノベーションの関係に展開することも十分可能であると考えられる。

### 2.3. 2 つの AI ガバナンス枠組み～Human-in-the-Loop (HITL) と Society-in-the-Loop (SITL) ～

AI のリスクガバナンスの責任については、AI 開発者や利用者にとって負わせることにより、コンプ

ライアンスコストが過大となり、イノベーションの制約となる可能性がある。リスクガバナンスとイノベーションを両立させるためには、AI 開発者や利用者と政府・制度の側の協力と責任分担が必要であり、技術と制度の共進化が求められている。

こうした観点から、各国の規制アプローチは、最終判断・責任を人間が担う、人間の介在 (HITL: Human in the Loop) に依拠して法的・倫理的責任を定義してきた。これにより、従来の責任分担ルールとの連続性を担保するとともに、人間の主体性(Agency)を確保することで、利用者の安心感をもたらすことができる (Fanni et al. (2023)、Legaspi et al. (2024))。具体的には、AI 利用者に対する異議申立権 (Contestability by Design) の付与 (Fanni et al. (2023)) や、透明性や人間の部分的な介入など人間の主体性の感覚を高めるインターフェースが信頼性向上のために重要である (Legaspi et al. (2024)) とされてきた。HITL の考え方は、AI の開発企業に対しても、人間を介在させる設計が必要という明確なシグナルを送ってきたと言える。

他方で、最近では、急速に自動化・自律化の方向へ進化している AI 技術の多様なリスクに対して、AI システム自体に社会的規範や法的ルールを内在化させる SITL (Society-in-the-Loop) のアプローチが注目されつつある (Rahwan (2018))。SITL では、AI の意思決定が社会全体の価値や倫理を反映するよう設計されることを目指すため、政策立案者、技術開発者、市民が協力し、社会全体の利益を考慮するよう適応的に AI を設計すべきとされる (Rahwan (2018))。このアプローチは責任分担ルールを革新することにより、初期段階では高度な技術開発を AI 開発者に求めることとなり、利用者の受容性も直ちには高まらないが、中長期的には、自律的 AI のイノベーションと社会受容性を支える基盤となる可能性がありうると考える。

#### 2.4. HITL/SITL 型 AI ガバナンスのイノベーション効果に対するポーター仮説の拡張

ポーター仮説は、規制が企業に新たな性能基準達成を促すことで、技術革新と競争力向上を同時に実現し得ることを示してきた。これを AI ガバナンス領域に拡張する際には、環境規制における単一の性能基準 (例: 排出削減) とは異なり、例えば、自動運転分野や医療 AI 分野において、安全性、説明可能性、社会的価値との整合性など多次元的な性能要求が同時に課される点の特徴である。さらに、AI 分野では規制がもたらす効果は技術的要求水準の提示だけでなく、企業や利用者の法的責任範囲の明確化や利用者・市民の社会的受容性向上にも直結している。この「制度シグナル」は、ポーター仮説に示されているとおり、不確実性の高い投資環境で企業の技術選択や研究開発方向を誘導し、投資リスクの軽減や市場失敗の是正を図ることができることに加え、AI 開発者や利用者の法的責任範囲の明確化と利用者の社会受容性を向上させることを通じて、イノベーション創発を促すことができると考えられる。

HITL (Human-in-the-Loop) と SITL (Society-in-the-Loop) は、異なる制度シグナルを発することで、企業の投資行動や技術選択に影響を与える。

- ✓ **HITL シグナル**: 人間介在を前提とした最終責任の明確化。企業にとって法的リスクを限定し、透明性や説明可能性の確保、利用者に安心感を与える UI/UX 改良などを促す。新規参入企業にとっては参入障壁の低下にもつながる。
- ✓ **SITL シグナル**: AI 自体に社会規範を内在化。初期コストは高いが、長期的にはシステムの自律的運用による先行者優位、他分野へのスピルオーバー効果を生み出す。

本研究では、これらの制度シグナルが企業イノベーションに作用する媒介経路を以下の3つに整理する。

##### ① リスク限定経路

企業の法的・経済的リスク認識を軽減し、資本コストの低下を通じて R&D 投資を促進する。

##### ② 信頼醸成経路

利用者・市民の信頼を高め、市場導入や実証実験の障壁を低減する。

##### ③ 需要創出経路

技術適合リストや実証実験など、承認・適合プロセスが市場に明確な技術要件を提示し、企業の技術探索行動を誘導する。

さらに、制度効果は技術ライフサイクルと相互作用する。新規市場形成期には HITL が短期的信頼と投資環境の安定化に寄与し、技術成熟期には SITL が長期的な社会受容性の確立に効果を発揮する。このため、HITL から SITL への段階的な規制トランジション設計が、持続的なイノベーションの促進に有効であると考えられる。

### 3. 分析枠組みと仮説

#### 3.1 分析の視点

本研究は、AI ガバナンスにおける Human in the Loop (HITL) 型と Society in the Loop (SITL) 型制度が、自動運転分野の技術革新（イノベーション）に与える影響を定量的に比較する。

特に、規制が企業に与える 3 つの「制度シグナル」（リスク限定、信頼醸成、需要創出）をポーター仮説の枠組みに位置づけ、制度シグナル → 企業行動変容 → イノベーション成果という 3 段階の因果経路を想定する。

なお、本研究における自動運転レベルの定義は、SAE International の J3016（2016 年 9 月）及びその日本語参考訳である JASO TP 180041（2018 年 2 月）の定義を採用する。

- ・レベル 3（条件付運転自動化）：システムが全ての動的運転タスクを限定領域において実行。作動継続が困難な場合は、運転者がシステムの介入要求等に適切に応答。安全運転に係る監視・対応主体は、システムだが、作動継続が困難な場合は運転者。
- ・レベル 4（高度運転自動化）：システムが全ての動的運転タスク及び作動継続が困難な場合への応答を限定領域において実行。安全運転に係る監視・対応主体は、システム。

#### 3.2 理論的背景

##### ● 弱いポーター仮説の拡張

弱いポーター仮説（適切な規制はイノベーションを促進する）を、AI ガバナンス分野に適用。HITL は短期的に市場受容性を高め、人間介在型の安全技術や透明性・説明可能性関連技術のイノベーションを促す。一方、SITL は長期的に社会規範の内在化を通じて自律性の高い AI のイノベーションを誘発する。

##### ● 制度間移行の動学

技術成熟度や市場受容度に応じ、HITL から SITL への段階的移行が有効な場合がある。

#### 3.3 仮説設定

##### ● 仮説 1（HITL 型規制のイノベーション創出効果）

運転者の最終的な安全運転義務を前提とするレベル 3 の自動運転など、利用者責任を求める度合いが高い分野では、HITL 型規制がイノベーション創出効果を高める。

##### ● 仮説 2（SITL 型規制のイノベーション創出効果）

自動運転システムを含む運行提供者側に責任を移管するレベル 4 自動運転など、利用者責任を求める度合いが低下し、システムの自律性が重視される分野では、HITL 型規制のイノベーション創出効果は乏しく、SITL 型規制が中長期的にイノベーション創出効果を高める。

##### ● 仮説 3（移行期の最適設計）

AI システムの自律化の程度（技術成熟度）が中程度の分野では、HITL から SITL への段階的移行が持続的なイノベーション成果をもたらす。

### 4. データと分析方法

#### 4.1 データ構造

本研究では、PATSTAT 2025 Spring データベースに基づき、自動運転制御システム（CPC: B60W60/001）の特許ファミリーを抽出した。分析対象期間は 2013～2023 年（11 年間）、上位 200 社を企業グループ別に名寄せした 87 グループを対象とする（出願件数が比較的少ない企業は除外した）。

説明変数として、ドイツおよび日本における自動運転関連法制の導入（年・国ダミー）を設定した。レベル 3（HITL 寄り）はドイツ 2017 年、日本 2019 年、レベル 4（SITL 寄り）は、ドイツ 2021 年、日本 2022 年である。加えて、各国における機械学習関連特許（G06N20）および画像処理関連特許（G06T）の年間出願件数を集計し、国際的な技術トレンドをコントロール変数として投入した。

#### 4.2 計量モデル

企業別パネルデータに対して、企業及び出願年の固定効果モデル（LSDV）を推定した。被説明変数は、出願件数の対数水準（ $\log(\text{Patent}+1)$ ）、出願件数の成長率（ $\Delta \log(\text{Patent}+1)$ ）の 2 種類を設定。

イベントスタディ設計として、自動運転関連法制の導入年を基準（ $t=0$ ）、導入前年（ $t=-1$ ）をレファレンス期とし、 $t \in \{-2, 0, +1, +2\}$  の係数を推定した（制度導入までの事前調整の効果を含めて分析）。

### 4.3 HITL/SITL 関連特許出願数の推計

6言語（英・日・独・中・韓・仏）対応の HITL/SITL キーワードを作成。このキーワードに合致するコアとなる HITL/SITL 特許出願（コア出願）を抽出。その上で、すべての特許出願について、コア出願との内容の近さを推計（センテンス BERT で文章をベクトル化し、ベクトルのコサイン距離で類似度を推計）。特許出願を類似度に応じて HITL、SITL、拮抗してどちらとも言いがたい（Tie）、どちらにも近くない（Far）に分類し、Tie/Far は Unclassified に分類。

## 5. 結果

### 5.1 記述統計

2013～2023年の累積で、自動運転制御システム関連特許出願件数（CPC: B60W60/001）は約2万1千件（HITL約6,300件、SITL約8,500件、未分類約6,300件）。上位200社の企業パネルの特許出願件数は約1万3千件。

### 5.2 イベントスタディ分析

#### 5.2.1 レベル3 法制的導入効果

ドイツ（2017年導入）では、導入直後から一貫して正の効果が確認され、特に導入2年後（ $k=+2$ ）に0.868（ $p=0.004$ ）と統計的に有意な上昇が見られた。Log-diffでも導入年・翌年に正の効果が観測され、HITL型制度が持続的に技術開発を促進したことが示唆される。

一方、日本（2019年導入）では、導入年（ $k=0$ ）に0.574（ $p=0.051$ ）、翌年（ $k=+1$ ）に0.769（ $p=0.021$ ）と一時的な増加が観測されたが、2年後（ $k=+2$ ）には効果が減衰した。短期的には出願促進効果がみられるが、持続性は限定的である。

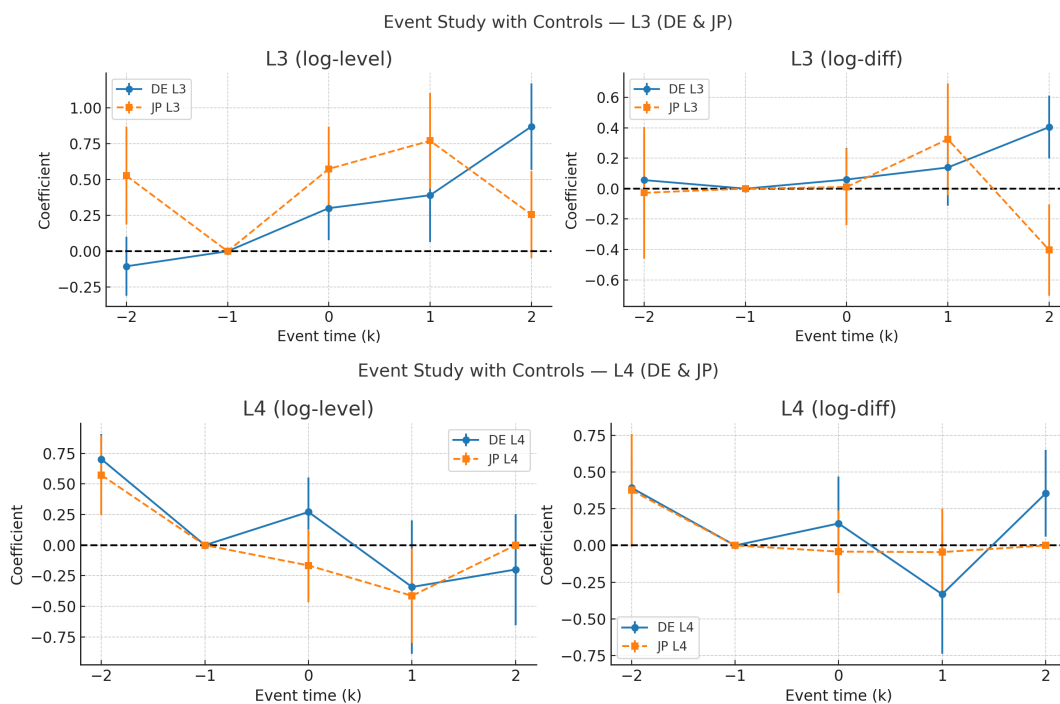
#### 5.2.2 レベル4 法制的導入効果

ドイツ（2021年導入）では、導入2年前（ $k=-2$ ）の正の係数が観測されたが、これは実際には2017年導入のL3の残存効果を反映していると解釈される。L4導入後（ $k=0, +1, +2$ ）の係数はいずれも有意ではなく、L3に比べ追加的なイノベーション効果は確認されなかった。

日本（2022年導入）については、データが2023年までであるため $k=+2$ は観測できない。導入年および翌年の効果は小さく、統計的有意性も確認されなかった。したがって、日本におけるL4独自の効果は限定的であった。

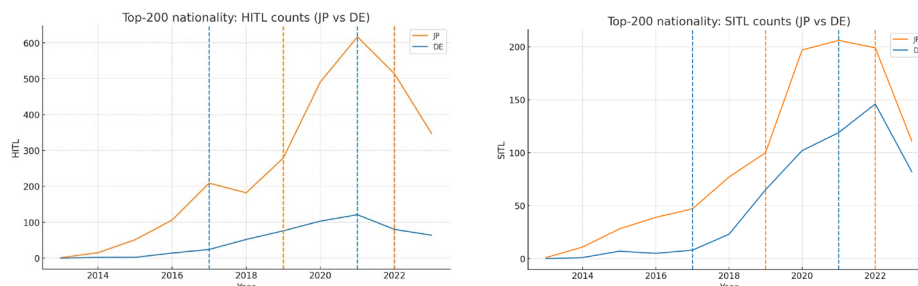
event	country	dep_var	-2	-1	0	1	2
L3	DE	log_Patent_Count_plus1	-0.105 (0.205, $p=0.609$ )	0 (baseline)	0.300 (0.225, $p=0.181$ )	0.390 (0.326, $p=0.232$ )	0.868 (0.301, $p=0.004$ )**
L3	JP	log_Patent_Count_plus1	0.526 (0.340, $p=0.122$ )	0 (baseline)	0.574 (0.293, $p=0.051$ )†	0.769 (0.334, $p=0.021$ )*	0.256 (0.303, $p=0.399$ )
L3	DE	dlog_Patent_Count_plus1	0.056 (0.294, $p=0.848$ )	0 (baseline)	0.060 (0.208, $p=0.773$ )	0.139 (0.253, $p=0.581$ )	0.405 (0.206, $p=0.050$ )*
L3	JP	dlog_Patent_Count_plus1	-0.027 (0.432, $p=0.950$ )	0 (baseline)	0.011 (0.252, $p=0.966$ )	0.326 (0.367, $p=0.376$ )	-0.404 (0.301, $p=0.180$ )
L4	DE	log_Patent_Count_plus1	0.703 (0.205, $p=0.001$ )**	0 (baseline)	0.271 (0.282, $p=0.336$ )	-0.342 (0.546, $p=0.531$ )	-0.199 (0.454, $p=0.661$ )
L4	JP	log_Patent_Count_plus1	0.573 (0.325, $p=0.078$ )†	0 (baseline)	-0.166 (0.300, $p=0.581$ )	-0.414 (0.386, $p=0.284$ )	NaN
L4	DE	dlog_Patent_Count_plus1	0.393 (0.202, $p=0.051$ )†	0 (baseline)	0.150 (0.321, $p=0.640$ )	-0.333 (0.406, $p=0.412$ )	0.356 (0.297, $p=0.231$ )
L4	JP	dlog_Patent_Count_plus1	0.377 (0.384, $p=0.326$ )	0 (baseline)	-0.042 (0.282, $p=0.882$ )	-0.045 (0.297, $p=0.879$ )	NaN

※数値は、係数（標準偏差、p値）。\*\*は1%有意、\*は5%有意、†は10%有意。



### 5.3 上位 200 企業の国籍別の HITL/SITL 関連特許出願の推移

- ドイツでは、SITL 関連特許出願数は L3 制度導入後（2018 年頃）から中期的に上昇。ピークは L4 制度導入直後の 2022 年。HITL 関連特許出願数は緩やかに上昇するも日本と比べると低い。
- 日本では、HITL 関連特許出願数は、L3 制度導入の 2 年前（2017 年）頃から上昇。ピークは L3 制度導入の 2 年後（2021 年）であり、L4 制度導入（2022 年）後は低下。SITL 関連出願数も L3 制度導入時（2019 年）頃から上昇するも、HITL 関連特許出願数の 1/3 程度の水準。



## 6. 考察と今後の課題

運転者の最終的な安全運転責任を前提とする HITL 型規制（L3）は、ドイツ企業及び日本企業において一定の特許出願総数の増加効果をもたらしたが、HITL/SITL 関連特許の区別まで分析すると、特に日本企業に対して顕著な HITL 関連特許出願増加をもたらした可能性がある。その効果は制度導入前から始まっているが、特に制度導入 2 年後にピークを迎えている。これは HITL 型の自動運転に関する利用者責任を明確化し、法的リスクを軽減することで企業投資を誘導する「リスク限定シグナル」として機能した可能性がある。ドイツ企業においても HITL 特許出願数の増加をもたらしたが、日本と比べると水準は低く、HITL 型の規制導入の効果は日本で特に有効であった可能性がある。

他方で、SITL 型規制（L4）は、ドイツ・日本ともに導入後の有意な効果は確認されなかった。これは、L3 による既存の上昇効果を上乗せできなかったためである。

本研究では、短期的には HITL によって市場受容性と投資誘因を確保し、中期以降は SITL に移行するという段階的設計が有効かどうかについて、特に、後段の SITL 規制の効果について十分なデータに基づく検証ができなかった。理由としては、SITL 関連特許の増加の少なさから、規制の性格自体が SITL 関連特許のイノベーション効果を持つに十分な条件を満たしていなかった可能性がある。また、SITL 分野の特許出願は CPC で B60W60/001 に限られない可能性もある。引き続き、時系列データを長期間とった上で、日本及びドイツ以外の企業の特許出願の傾向の分析を含め、詳細な調査・分析が必要である。

## 参考文献

- Ambec, S., Cohen, M. A., Elgie, S., & Lanoie, P. (2013). The porter hypothesis at 20: Can environmental regulation enhance innovation and competitiveness? *Review of Environmental Economics and Policy*, 7(1), 2–22. <https://doi.org/10.1093/reep/res016>
- Blind, K., Petersen, S. S., & Riillo, C. A. F. (2017). The impact of standards and regulation on innovation in uncertain markets. *Research Policy*, 46(1), 249–264. <https://doi.org/10.1016/j.respol.2016.11.003>
- Fanni, R., Steinkogler, V. E., Zampedri, G., & Pierson, J. (2023). Enhancing human agency through redress in Artificial Intelligence Systems. *AI and Society*, 38(2), 537–547. <https://doi.org/10.1007/s00146-022-01454-7>
- Legaspi, R., Xu, W., Konishi, T., Wada, S., Kobayashi, N., Naruse, Y., & Ishikawa, Y. (2024). The sense of agency in human-AI interactions. *Knowledge-Based Systems*, 286. <https://doi.org/10.1016/j.knosys.2023.111298>
- Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14. <https://doi.org/10.1007/s10676-017-9430-8>
- Porter, M. E., & van der Linde, C. (1995). Toward a New Conception of the Environment-Competitiveness Relationship. In *Journal of Economic Perspectives* (Vol. 9, Issue 4).
- 市川類 (2023) 「技術と制度の共進化の中での地域的多様性による制度イノベーションの進展～世界の人工知能 (AI) ガバナンス制度の進化メカニズム」 IIR Working paper WP#23-01
- 高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議 (2018) 「自動運転に係る制度整備大綱」
- 谷川浩也 (2003) 「環境問題と技術革新—ポーター仮説の今日的意義—」 経済産業ジャーナル (独立行政法人経済産業研究所)